



## HemaExplorer

### a database of mRNA expression profiles in normal and malignant haematopoiesis

Bagger, Frederik Otzen; Rapin, Nicolas Philippe Jean-Pierre; Theilgaard-Mønch, Kim; Kaczkowski, Bogumil; Thorén, Lina Anna Maria; Jendholm, Lars Johan; Winther, Ole; Porse, Bo Torben

*Published in:*  
Nucleic Acids Research

*DOI:*  
[10.1093/nar/gks1021](https://doi.org/10.1093/nar/gks1021)

*Publication date:*  
2013

*Document version*  
Publisher's PDF, also known as Version of record

*Citation for published version (APA):*  
Bagger, F. O., Rapin, N. P. J-P., Theilgaard-Mønch, K., Kaczkowski, B., Thorén, L. A. M., Jendholm, L. J., Winther, O., & Porse, B. T. (2013). HemaExplorer: a database of mRNA expression profiles in normal and malignant haematopoiesis. *Nucleic Acids Research*, 41(Database issue), D1034-D1039.  
<https://doi.org/10.1093/nar/gks1021>

# HemaExplorer: a database of mRNA expression profiles in normal and malignant haematopoiesis

Frederik Otzen Bagger<sup>1,2,3,4</sup>, Nicolas Rapin<sup>1,2,3,4,\*</sup>, Kim Theilgaard-Mönch<sup>2,3,4,5</sup>, Bogumil Kaczkowski<sup>1,3</sup>, Lina A. Thoren<sup>2,3,4</sup>, Johan Jendholm<sup>2,3,4</sup>, Ole Winther<sup>1,3,6,\*</sup> and Bo T. Porse<sup>2,3,4,\*</sup>

<sup>1</sup>Bioinformatics Centre, Department of Biology, <sup>2</sup>The Finsen Laboratory, Rigshospitalet, Faculty of Health Sciences, <sup>3</sup>Biotech Research and Innovation Center (BRIC), <sup>4</sup>Danish Stem Cell Centre (DanStem), Faculty of Health Sciences, University of Copenhagen, Copenhagen, DK2200 Denmark, <sup>5</sup>Dept. of Hematology, Skånes University Hospital, University of Lund, 221 85 Lund, Sweden and <sup>6</sup>Intelligent Signal Processing, Informatics and Mathematical Modelling (IMM), Technical University of Denmark (DTU), DK2800 Kgs. Lyngby, Denmark

Received August 14, 2012; Revised October 2, 2012; Accepted October 3, 2012

## ABSTRACT

The HemaExplorer (<http://servers.binf.ku.dk/hema-explorer>) is a curated database of processed mRNA Gene expression profiles (GEPs) that provides an easy display of gene expression in haematopoietic cells. HemaExplorer contains GEPs derived from mouse/human haematopoietic stem and progenitor cells as well as from more differentiated cell types. Moreover, data from distinct subtypes of human acute myeloid leukemia is included in the database allowing researchers to directly compare gene expression of leukemic cells with those of their closest normal counterpart. Normalization and batch correction lead to full integrity of the data in the database. The HemaExplorer has comprehensive visualization interface that can make it useful as a daily tool for biologists and cancer researchers to assess the expression patterns of genes encountered in research or literature. HemaExplorer is relevant for all research within the fields of leukemia, immunology, cell differentiation and the biology of the haematopoietic system.

## INTRODUCTION

Haematopoiesis is the developmental process by which all blood cells are formed. The haematopoietic system is organized in a hierarchical manner with the haematopoietic stem cell (HSC) residing at the apex (Figure 1). HSCs have the ability to self-renew for life and to

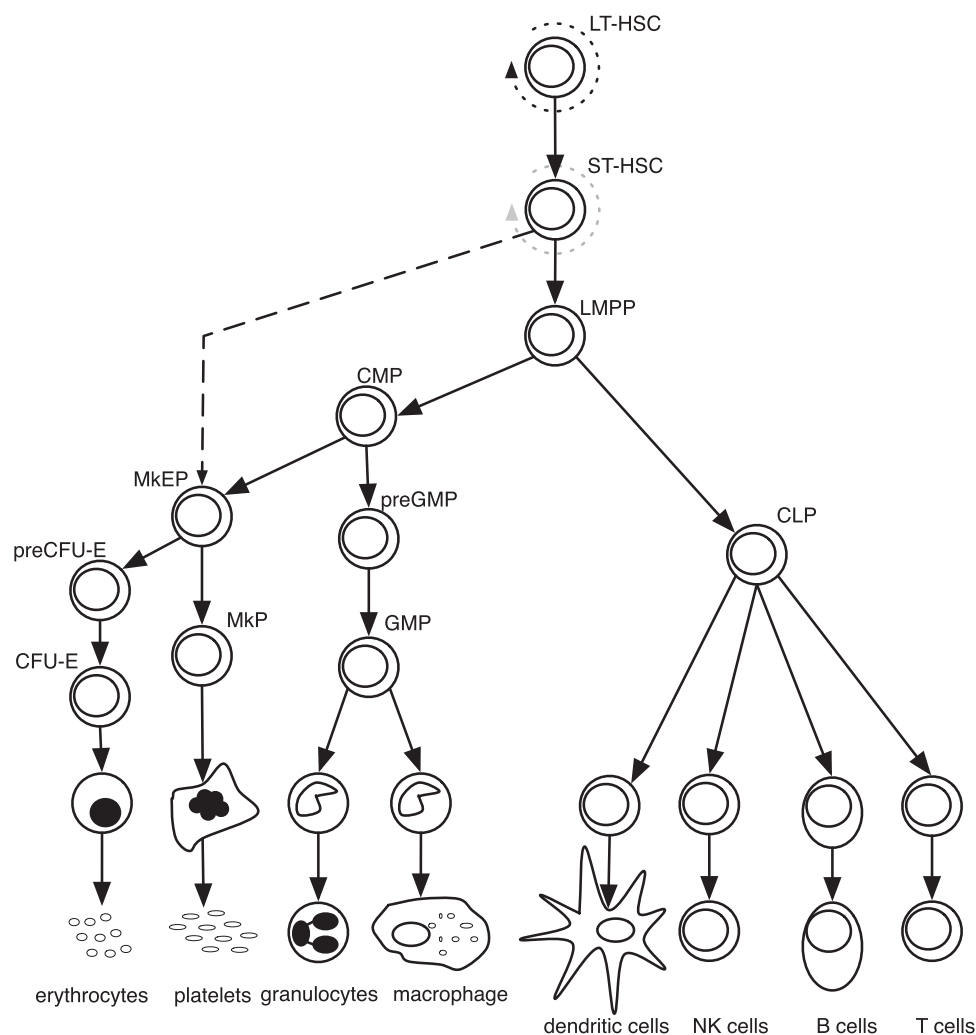
differentiate into intermediate progenitor cells that subsequently generate the mature cells of the blood stream and immune system (3). Thus, HSCs are ultimately responsible for the constant and life-long generation of intermediate progenitors and terminally differentiated cells. Genetic and epigenetic aberrations may lead to a block in normal haematopoietic differentiation, which may result in the development of Acute Myeloid Leukemia (AML), which is an aggressive blood cancer.

Given the general interest in haematopoiesis and its associated malignancies such as AML, a number of databases now addresses questions related to these fields, a majority of which is focused on mRNA expression data. There are multiple tools available online that provides a quick overview of the expression of single genes in a selection of tissues as well as some sorted cells. The requirement for purified sorted cell populations in the study of haematopoiesis makes the curation of the database as well as the availability of a reasonable selection of haematopoietic cells important. Examples of general and non-curated databases with focus on representing gene expression based on mRNA microarrays are Gene expression Atlas (GXA) (4), GeneNote (5) and BioGPS (6).

GXA provides a count table of over and under expression of a query gene in a wide range of cell types and experimental setups, presently counting 3476 experiments of various types. The count numbers are based on differential expression within each microarray dataset in a selection of publicly available microarray data. GeneNote provides a bar chart with a continuous ratio scale axis, displaying expression intensities pre-processed and scaled to a common mean in a global normalization. The database

\*To whom correspondence should be addressed. Tel: +45 3545 6023; Fax: +45 3545 3797; Email: bo.porse@finsenlab.dk  
Correspondence may also be addressed to Nicolas Rapin. Tel: +45 24 43 28 78; Fax: +45 3545 3797; Email: nicolas.rapin@bric.ku.dk  
Correspondence may also be addressed to Ole Winther. Tel: +45 4525 3895; Fax: +45 4587 2599; Email: owi@imm.dtu.dk

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.



**Figure 1.** A model of murine haematopoiesis (1). Dotted line represents self-renewal. Dashed line indicates the separation of M&E lineage from the lymphoid and the remaining myeloid populations as shown by Adolfsson *et al.* (2). *Abbreviations are:* LT-HSC, long-term haematopoietic stem cell; ST-HSC, short-term haematopoietic stem cell; LMPP, lymphoid-primed multipotential progenitors; CLP, common lymphoid progenitor cells; CMP, common myeloid progenitor cell; MEP, megakaryocyte-erythroid progenitor cell; preGMP, pre-granulocyte monocyte progenitors; GMP, granulocyte monocyte progenitors; M&EP, megakaryocyte erythroid precursors; MkP, megakaryocyte precursor; PreCFU-E, pre-colony-forming unit erythroid cells; CFU-E, colony-forming unit erythroid cells; NK cells, natural killer cells.

contains a coherent collection of expression profiles in 28 healthy human tissues types. BioGPS provides a summary of single genes including a chart of gene expression based on a single microarray dataset. The variety of cell or tissue types depends on the selected dataset.

Databases dedicated to research in the haematopoietic system and haematopoietic disorders includes Leukemia Gene Atlas (7), BloodExpress (8) and Haematopoietic Fingerprint (9). The Leukemia Gene atlas comprises 20 datasets from Leukemia research experiments; 13 of those sets are gene expression profiles (GEPs). The database provides the possibility of running a number of standardized gene expression analysis methods on a selected dataset, including value sorted bar plots of expression.

BloodExplorer is a database of 271 microarray-based experiments and provides a list of datasets where a query gene is 'present' in terms of a significant *P*-value and presence on the microarray chip.

Haematopoietic Fingerprints database contains data from a mouse dataset with 20 samples, and can provide bar charts of gene expression in haematopoietic stem cells and mature populations of erythrocytes, granulocytes, monocytes, NK cells, activated and naive T cells, and B cells. The database also provides an overview of expression across the chromosomes for each cell type.

The time and expenses necessary for sorting cells drastically reduces the availability of data for investigation of haematopoiesis, as compared to whole tissue samples. This is particularly the case for stem and early progenitor cell populations. None of the present databases integrates and compare GEPs from haematopoietic cells derived from several labs and experiments. Furthermore, there are no publicly available databases with the possibility of visualizing gene expression in a range of sorted cells from both human and murine haematopoietic system, just like a direct comparison with human AML in such

a normal context presently will require extensive computational or experimental work from the researcher.

Here, we present the first publically available fully integrated and curated database of GEPs of sorted cells from stem, progenitor and mature populations important for haematopoietic development in both mouse and human. Compared to the present databases, this method provides a far more complete view of expression in the haematopoietic system. The interface allows for easy visualization of the expression of single genes as well as tools to assess correlation of two genes across multiple cell types. Furthermore the database includes bone marrow from four karyotypic distinct AML subtypes and provides two options for investigating malignant haematopoiesis compared to the normal context.

## MATERIALS AND METHODS

The database is built from a curated selection of publicly available datasets containing GEPs from sorted cell populations. The database integrates several Affymetrix platforms (HG-U133A, HG-U133B and HG-U133 Plus 2; Mouse Genome 430 2.0). For the human data the overlapping probes across the platforms populate the database. The mouse database is solely based on the Mouse Genome 430 2.0 platform. Data were separated into batches divided by platform and laboratory. The batches were normalized by robust multi-array average (RMA) using the affy package for R (10) and corrected for batch effects using ComBat (11). ComBat is an empirical Bayes (EB) method that: First, standardizes the data by a gene-wise least-square. Secondly, estimates EB batch effect parameters by the method of moments using Gaussian and inverse gamma as prior distributions. Lastly, the data are adjusted using the calculated batch effect estimators. This procedure leaves all data in the database directly comparable between samples within the human and murine part of the database, respectively.

From the normalized and batch corrected data a new dataset was build containing fold changes between human AML and the closest normal counterpart, as identified by projecting data to a reduced space defined by principle component analysis and deciding Euclidian distance by *k*-nearest neighbour method. (Rapin *et al.*, manuscript in preparation). The model parameters are constant for the present set of normal sorted cells, and addition of future AMLs will not influence the existing samples.

In order to handle gene aliases a dictionary of gene aliases was constructed from NCBI {ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/} and The HUGO Gene Nomenclature Committee (HGNC) {www.genenames.org}. Ambiguous gene aliases were not included when constructing the dictionary.

The database structure is a combination of Python pickle and HDF5 files. The main data handling and graphic construction were written in Python NumPy and R called from Python via the rpy2 interface (12). Figure 2 shows a diagram of the database structure and query handling of single gene (*ERG*) lookup (13).

The database is continuously updated with new data as it becomes available. The present selection has been

focused on including as little batch effect as possible, and thus introducing as few platforms and data sources as possible.

## RESULTS AND DISCUSSION

The HemaExplorer has an easy-to-use interface with a single query field and the possibility to select between several data sources. At present, the database has available: human normal hematopoietic system, human AML, human normal hematopoietic system + human AML, human AML compared with closest normal counterpart, and mouse haematopoietic system. Cell types and data source are listed in Table 1. The database provides plot of expression for query genes and accepts unambiguous gene aliases. In the case of ambiguous gene aliases a deep link to genecards.org (24) with such a query is constructed on the interface level, to allow the user to identify the relevant gene symbol.

Currently the interface allows visualization of expression of single genes in haematopoietic cells and any two genes can also be plotted in a pair-wise correlation plot. There is a possibility to include expression in bulk bone marrow of four karyotypes of AML: AML with t(8;21), AML with t(15;17), AML with inv(16)/t(16;16) and AML with t(11q23)/MLL. Furthermore, the interrelationship between these four AML karyotypes can be investigated in a plot displaying their fold change with respect to the closest normal counterpart. In an advanced mode the user can select which of the available haematopoietic cell types to be displayed in the output plot, be it single genes, or two genes for a correlation plot.

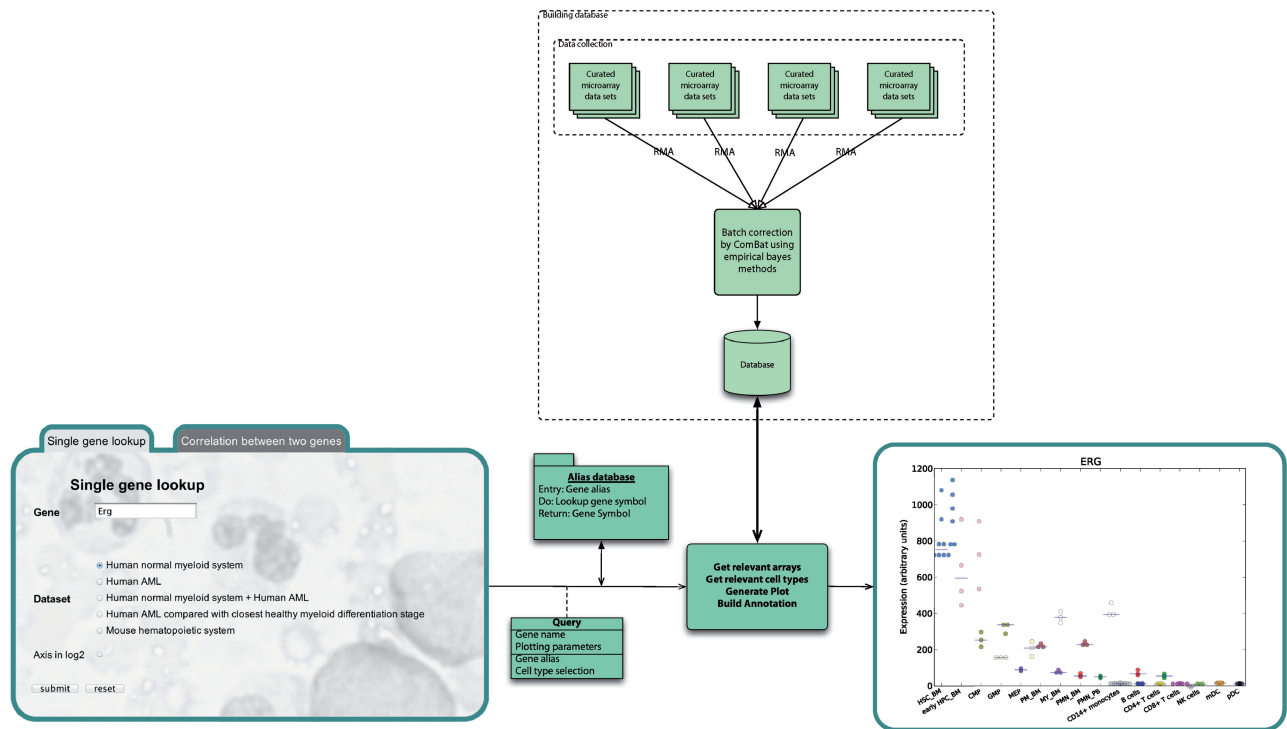
### Database

The overall integrity of the database is confirmed by dimensionality reduction technique (principle component analysis), as well as correlation analysis of expression profiles by hierarchical clustering, visualized in heatmaps (Supplementary Figure S1 and S2). Both validation techniques show a separation of cell types into well-defined clusters, and reveal no remaining batch effects. The plots are generated using the probe set with the highest mean expression, but normalized expression for all probes annotated to the query gene can be downloaded as a flat file on the result page.

### Example of correlation plot between two genes

The possibility of visualizing correlation between two genes makes it possible to identify genes that show similar expression across cell types. Furthermore, this feature enables researchers to visualize characterizing gene pairs with expression patterns that may separate cell populations into distinct clusters and could therefore potentially improve cell sorting strategies.

In Figure 3, an example of a FACS analysis for murine haematopoietic cells is depicted together with expression correlation plots from HemaExplorer. Gene expression in the correlation plots corresponds to the gene that encodes the surface marker used to gate cell populations in the FACS sorting strategy, performed according to Pronk



**Figure 2.** Diagram of database construction and output from query. *Top-down:* The database is build from publicly available data, normalized by RMA and batch corrected by ComBat (11). *Right-to-left:* Queries are sent to gene alias directory, and relevant gene expression data are retrieved from the repository. Plot is generated from gene expression data together with figure text and list of abbreviations used in the plot and formatted in to a result page.

**Table 1.** Cell types and sources of data currently in the HemaExplorer database

Samples	Source	Acc. number	ref
MY, pbPMN, bmPMN, PM	GEO	GSE19556	(14)
Monocytes	Array express	E-MEXP-1242	(15)
HSC BM, early HPC BM	GEO	GSE2666	(16)
CMP, GMP, MEP	GEO	GSE19599	(17)
HSC BM	GEO	GSE17054	(18)
Monocytes	GEO	GSE11864	(19)
Monocytes, B cells, CD4+, NK, CD8+, Eosinophils, mDC, Neutrophils, pDC	GEO	GSE28490	(20)
Leukemia samples	GEO	GSE13159	(21,22)
CD4+, CFU-E, CLP, ETP, GM, preGMP, IgM+SP, LMPP, LTHSC, MkE, MkP, NK, PreB, PreCFU-E, ProB, ProE, ST-HSC	GEO	GSE14833	(23)
NK, CD8+ naive, CD4+ act., CD8+ act., B-Cell, Monocytes, Granulocytes, Nucl. Erythrocytes	GEO	GSE6506	(7)

Grey-shaded cells contain mouse data and non-shaded cells contain human data.

**Abbreviations are:** MY, myelocytes; pbPMN, polymorphonuclear cells from peripheral blood; bmPMN, polymorphonuclear cells from bone marrow; PM, promyelocytes; HSC BM, haematopoietic stem cells from bone marrow; early HPC BM, early haematopoietic progenitor cells from bone marrow; CMP, common myeloid progenitor cells; GMP, granulocyte macrophage progenitor cells; MEP, megakaryocyte-erythroid progenitor cells; CD4+, CD4 positive T-cells; CD8+, CD8 positive T-cells; mDC, myeloid dendritic cells; pDC, plasmacytoid dendritic cells; CFU-E, colony-forming unit erythroid cells; CLP, common lymphoid progenitor cells; ETP, early thymic progenitor cells; preGM, pre granulocyte monocyte progenitor cells; GMP, granulocyte monocyte progenitor cells; IgM+SP, IgM positive spleen cells; LMPP, lymphoid-primed multipotential progenitor cells; LT-HSC, long-term haematopoietic stem cells; MkE, megakaryocyte erythroid cells; MkP, megakaryocyte precursor cells; NK, natural killer cells; preB, pre-B cells; PreCFU-E, pre-colony-forming unit erythroid cells; ProB, B-cell progenitor cells; ProE, erythroid progenitor cells; ST-HSC, short-term haematopoietic stem cell.

*et al.* (1). FACS data and gates in Figure 3A and B represent external biological samples from which there are no expression data in the database, and serves illustrative purposes only.

The comparison between a FACS sorting strategy and the expression correlation plots from HemaExplorer shows similar clustering of cell types.

### Example of single gene lookup

In Figure 3E, a single gene lookup for *BMI-1* is depicted. The *BMI-1* gene is known for its role in proliferation, senescence and self-renewal under various conditions (25–27). The expression plot for *BMI-1* shows high expression for immature cells, gradually decreasing for cell types



found in later stages of the myeloid haematopoietic differentiation pathway. *BMI-1* expression in the four AML karyotypes included in the database has levels comparable to the more immature stages, suggesting that the bulk of cells in leukemic bone marrow tends to be less mature and expresses more stem-like phenotype.

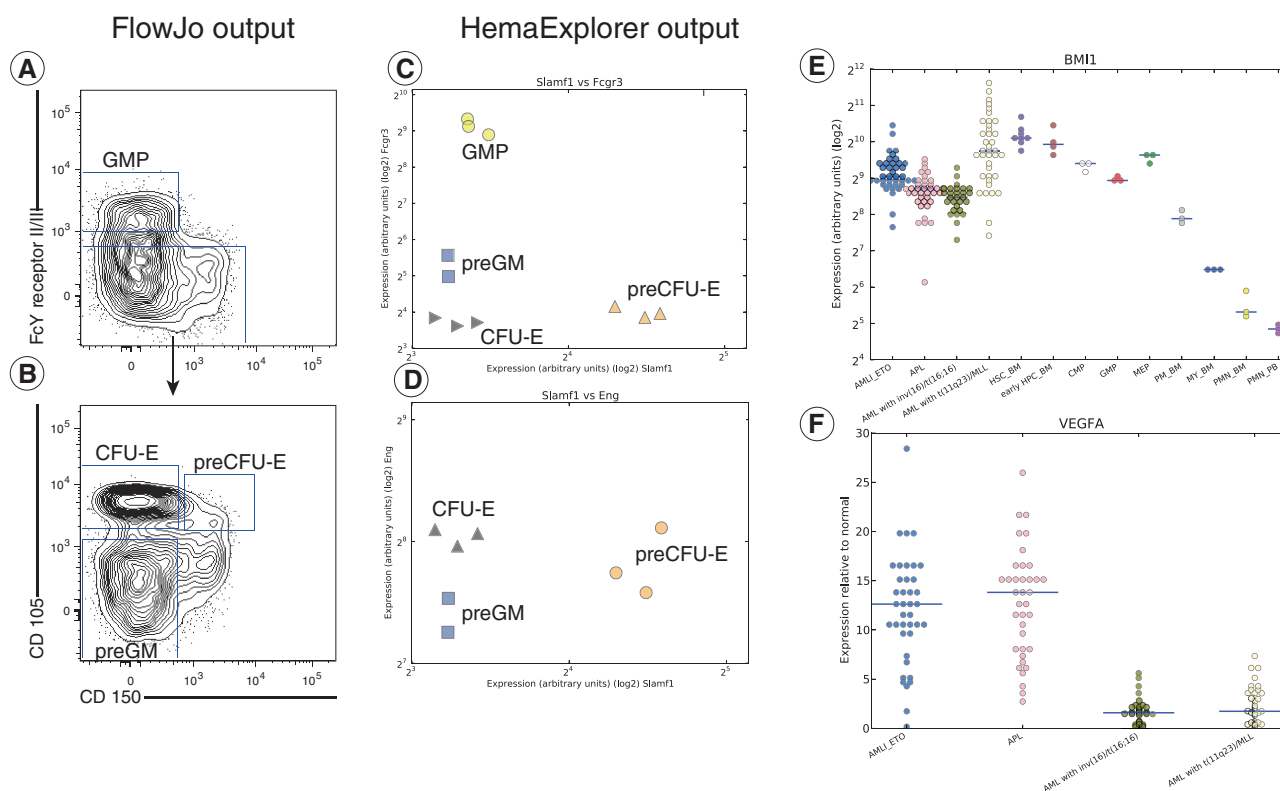
#### Example of comparison between human AML and the closest normal counterpart

In Figure 3F, a single gene lookup is depicted for the query gene *VEGFA* (vascular endothelial growth factor A). The plot shows the fold change between the AML and the closest normal counterpart. The plot shows clear difference in mRNA expression of *VEGFA* between the two groups: AML with t(8;21) (AML1\_ETO) and AML with t(15;17) (APL) on one side and AML with inv(16)/t(16;16) and AML with t(11q23)/MLL on the other side. This, and similar distinct separations of AML karyotypes when they are compared to their closest healthy counterpart, can provide researchers with valuable research targets then investigating differences between gene expression in discrete AML subtypes relative to normal cells.

In conclusion, the HemaExplorer is a curated normalized and batch corrected database of GEPs in normal and malignant haematopoiesis in human and mouse. The easy-to-use interface allows for an easy lookup of the expression levels of a gene and for the correlation of expression between pairs of genes. Full integration and comparability of data collected from several sources extends the scope, and possible overview of single mRNA expression in haematopoiesis, compared to the present public available databases. Moreover, the HemaExplorer contains four karyotypes of human AML that can be put in context of the normal haematopoietic system, when visualizing gene expression. Therefore, HemaExplorer will be of use to researchers within the fields of leukemia, immunology, cell differentiation and the biology of the haematopoietic system as a powerful easy-accessible tool for the assessment of gene expression.

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1 and 2.



**Figure 3.** (A–D) FACS analysis of a subpopulation of murine bone marrow cells depicted together with HemaExplorer gene expression correlation plots. Correlation between gene expression is shown for genes encoding the markers used in the FACS plot shown to the right. (A) FACS gating strategy for separating GMP from CFU-E, preCFU-E and preGM cells. (B) FACS gating strategy for separating CFU-E, preCFU-E and preGM cells. (E) Single gene lookup in the HemaExplorer of the expression of *BMI-1* in AML normal sorted cell populations from the myeloid compartment. (F) Single gene lookup in the HemaExplorer of the *VEGFA* gene. The dataset is human AML compared with the closest healthy myeloid differentiation stage and presents bone marrow samples from AML patients. Abbreviations are: HSC\_BM, haematopoietic stem cells from bone marrow; early HPC\_BM, haematopoietic progenitor cells from bone marrow; CMP, common myeloid progenitor cell; GMP, granulocyte monocyte progenitors; MEP, megakaryocyte-erythroid progenitor cell; PM\_BM, promyelocyte from bone marrow; MY\_BM, myelocyte from bone marrow; PMN\_BM, polymorphonuclear cells from bone marrow; PMN\_PB, polymorphonuclear cells from peripheral blood; AML1\_ETO, AML with t(8;21); APL, AML with t(15;17).

## FUNDING

Danish Research Council for Strategic Research; NovoNordisk Foundation. Funding for open access charge: Rigshospitalet.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Pronk, C.J.H., Rossi, D.J., Månsson, R., Attema, J.L., Norddahl, G.L., Chan, C.K.F., Sigvardsson, M., Weissman, I.L. and Bryder, D. (2007) Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. *Cell Stem Cell*, **1**, 428–442.
2. Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.-J., Thoren, L.A.M. *et al.* (2005) Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell*, **121**, 295–306.
3. Orkin, S.H. and Zon, L.I. (2008) Hematopoiesis: an evolving paradigm for stem cell biology. *Cell*, **132**, 631–644.
4. Kapushesky, M., Adamusiak, T., Burdett, T., Culhane, A., Farne, A., Filippov, A., Holloway, E., Klebanov, A., Kryvych, N., Kurbatova, N. *et al.* (2012) Gene Expression Atlas update—a value-added database of microarray and sequencing-based functional genomics experiments. *Nucleic Acids Res.*, **40**, D1077–D1081.
5. Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E. *et al.* (2005) Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, **21**, 650–659.
6. Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C.L., Haase, J., Janes, J., Huss, J.W. *et al.* (2009) BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.*, **10**, R130.
7. Hebestreit, K., Gröttrup, S., Emden, D., Veerkamp, J., Ruckert, C., Klein, H.-U., Müller-Tidow, C. and Dugas, M. (2012) Leukemia gene atlas – a public platform for integrative exploration of genome-wide molecular data. *PLoS One*, **7**, e39148.
8. Miranda-Saavedra, D., De, S., Trotter, M.W., Teichmann, S.A. and Götgens, B. (2009) BloodExpress: a database of gene expression in mouse haematopoiesis. *Nucleic Acids Res.*, **37**, D873–D879.
9. Chambers, S.M., Boles, N.C., Lin, K.K., Tierney, M.P., Bowman, T.V., Bradfute, S.B., Chen, A.J., Merchant, A.A., Sirin, O., Weksberg, D.C. *et al.* (2007) Hematopoietic fingerprints: an expression database of stem cells and their progeny. *Cell Stem Cell*, **1**, 578–591.
10. Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U. and Speed, T.P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.
11. Johnson, W.E., Li, C. and Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, **8**, 118–127.
12. Gautier, L. (2010) An intuitive Python interface for Bioconductor libraries demonstrates the utility of language translators. *BMC Bioinform.*, **11**(Suppl. 1), S11.
13. Bagger, F.O., Rapin, N., Theilgaard-Mönch, K., Kaczowski, B., Jendholm, J., Winther, O. and Porse, B. (2012) HemaExplorer: a Web server for easy and fast visualization of gene expression in normal and malignant hematopoiesis. *Blood*, **119**, 6394–6395.
14. Marstrand, T.T., Borup, R., Willer, A., Borregaard, N., Sandelin, A., Porse, B.T. and Theilgaard-Mönch, K. (2010) A conceptual framework for the identification of candidate drugs and drug targets in acute promyelocytic leukemia. *Leukemia*, **24**, 1265–1275.
15. Wildenberg, M.E., van Helden-Meeuwsen, C.G., van de Merwe, J.P., Drexhage, H.A. and Versnel, M.A. (2008) Systemic increase in type I interferon activity in Sjögren's syndrome: a putative role for plasmacytoid dendritic cells. *Eur. J. Immunol.*, **38**, 2024–2033.
16. Eckfeldt, C.E., Mendenhall, E.M., Flynn, C.M., Wang, T.-F., Pickart, M.A., Grindle, S.M., Ekker, S.C. and Verfaillie, C.M. (2005) Functional analysis of human hematopoietic stem cell gene expression using zebrafish. *PLoS Biol.*, **3**, e254.
17. Andersson, A., Edén, P., Olofsson, T. and Fioretos, T. (2010) Gene expression signatures in childhood acute leukemias are largely unique and distinct from those of normal tissues and other malignancies. *BMC Med. Genom.*, **3**, 6.
18. Majeti, R., Becker, M.W., Tian, Q., Lee, T.-L.M., Yan, X., Liu, R., Chiang, J.-H., Hood, L., Clarke, M.F. and Weissman, I.L. (2009) Dysregulated gene expression networks in human acute myelogenous leukemia stem cells. *Proc. Natl Acad. Sci. USA*, **106**, 3396–3401.
19. Hu, X., Chung, A.Y., Wu, I., Foldi, J., Chen, J., Ji, J.D., Tateya, T., Kang, Y.J., Han, J., Gessler, M. *et al.* (2008) Integrated regulation of Toll-like receptor responses by Notch and interferon-gamma pathways. *Immunity*, **29**, 691–703.
20. Allantaz, F., Cheng, D.T., Bergauer, T., Ravindran, P., Rossier, M.F., Ebeling, M., Badi, L., Reis, B., Bitter, H., D'Asaro, M. *et al.* (2012) Expression profiling of human immune cell subsets identifies miRNA-mRNA regulatory relationships correlated with cell type specific expression. *PLoS One*, **7**, e29979.
21. Kohlmann, A., Kipps, T.J., Rassenti, L.Z., Downing, J.R., Shurtleff, S.A., Mills, K.I., Gilkes, A.F., Hofmann, W.-K., Basso, G., Dell'orto, M.C. *et al.* (2008) An international standardization programme towards the application of gene expression profiling in routine leukaemia diagnostics: the Microarray Innovations in LEukemia study prephase. *Br. J. Haematol.*, **142**, 802–807.
22. Haferlach, T., Kohlmann, A., Wiczorek, L., Basso, G., Kronnie, G.T., Béné, M.-C., De Vos, J., Hernández, J.M., Hofmann, W.-K., Mills, K.I. *et al.* (2010) Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the International Microarray Innovations in Leukemia Study Group. *J. Clin. Oncol.*, **28**, 2529–2537.
23. Di Tullio, A., Vu Manh, T.P., Schubert, A., Månsson, R. and Graf, T. (2011) CCAAT/enhancer binding protein alpha (C/EBP(alpha))-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. *Proc. Natl Acad. Sci. USA*, **108**, 17016–17021.
24. Safran, M., Dalah, I., Alexander, J., Rosen, N., Iny Stein, T., Shmoish, M., Nativ, N., Bahir, I., Doniger, T., Krug, H. *et al.* (2010) GeneCards Version 3: the human gene integrator. *Database*, **2010**, baq020.
25. Lessard, J. and Sauvageau, G. (2003) Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature*, **423**, 255–260.
26. Park, I., Qian, D., Kiel, M., Becker, M.W., Pihlaja, M., Weissman, I.L., Morrison, S.J. and Clarke, M.F. (2003) Bmi-1 is required for maintenance of adult self-renewing hematopoietic stem cells. *Nature*, **423**, 302–305.
27. Wang, Y., Guan, Y., Wang, F., Huang, A., Wang, S. and Zhang, Y.A. (2010) Bmi-1 regulates self-renewal, proliferation and senescence of human fetal neural stem cells in vitro. *Neurosci. Lett.*, **476**, 74–78.